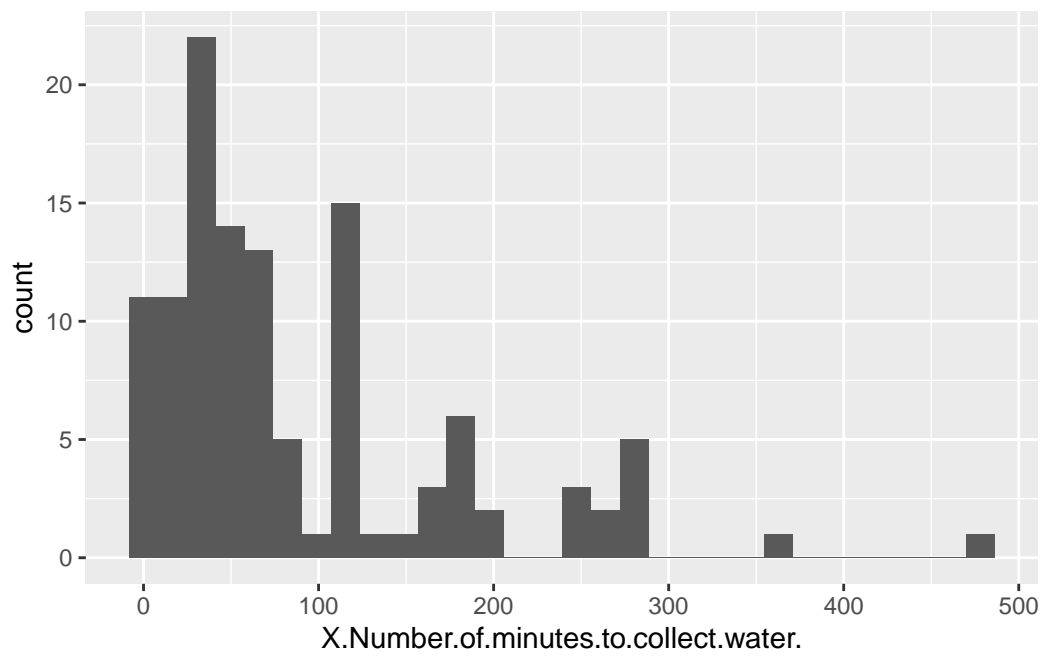
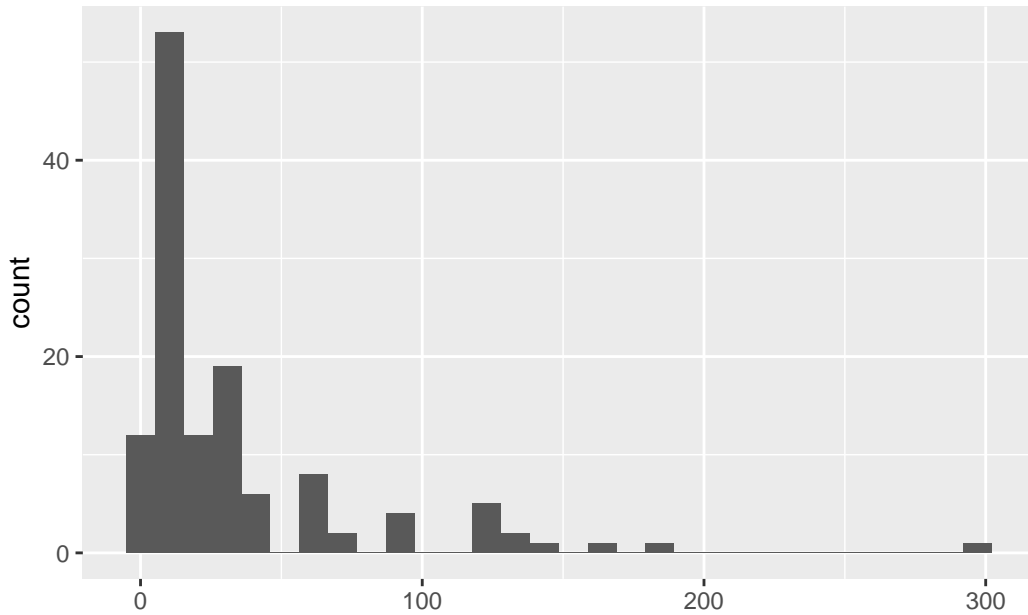


# Water for Good

## Report

### Uni-variate EDA



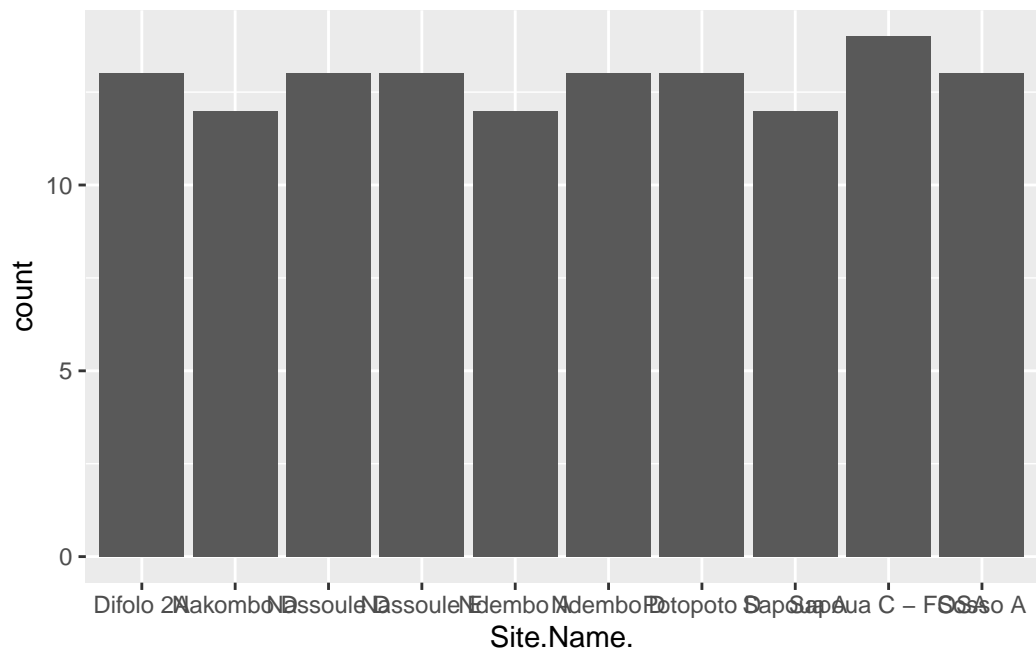
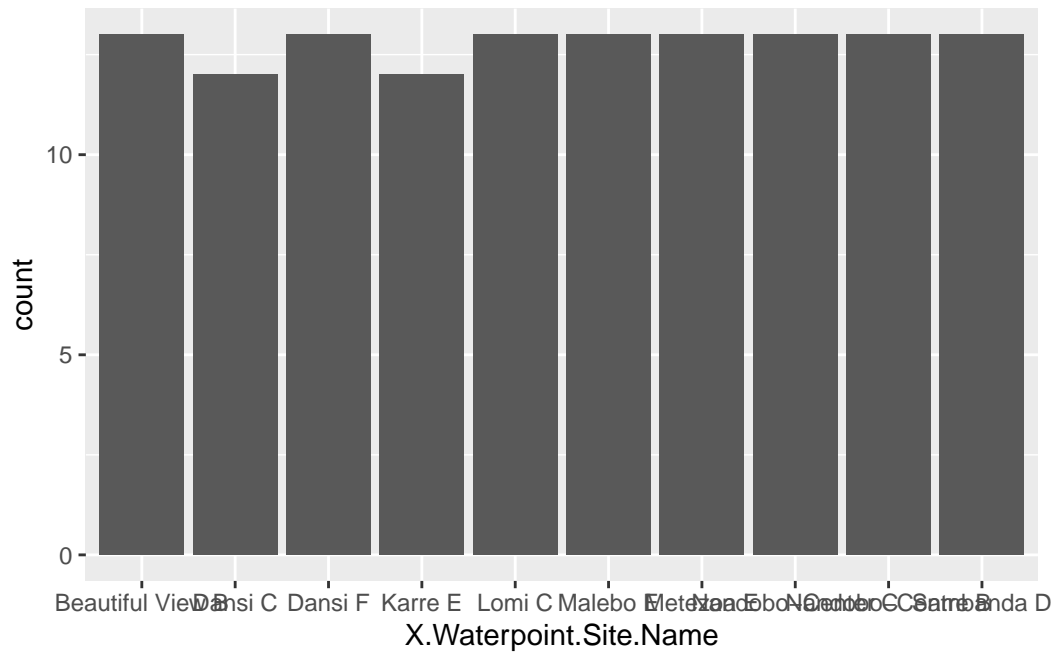


minutes.it.takes.to.walk.from.the.house.to.the.water.point..collect.the.water..including.v

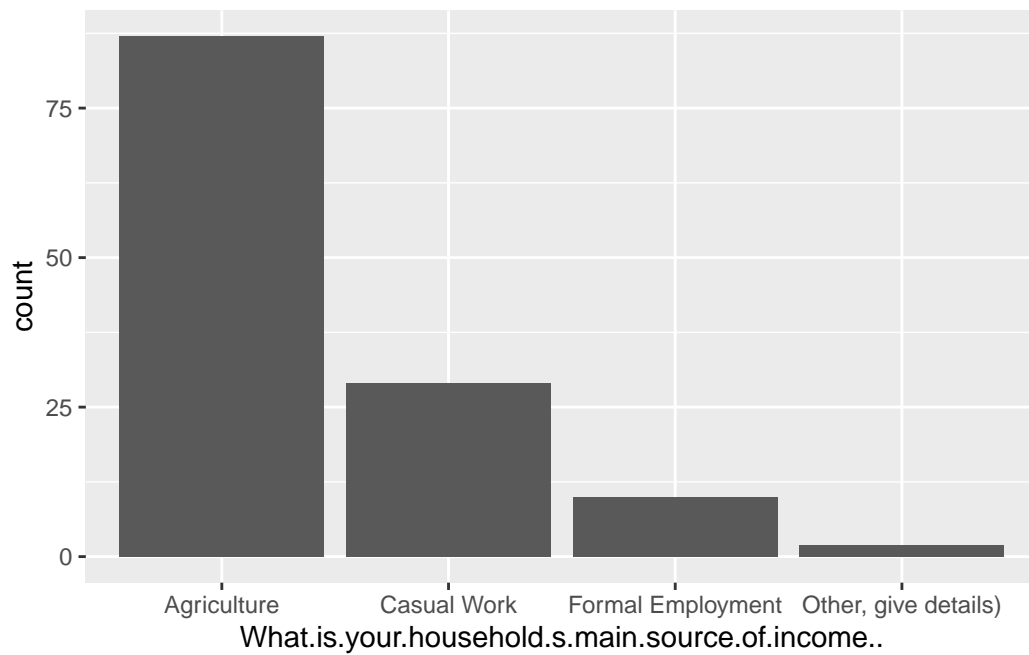
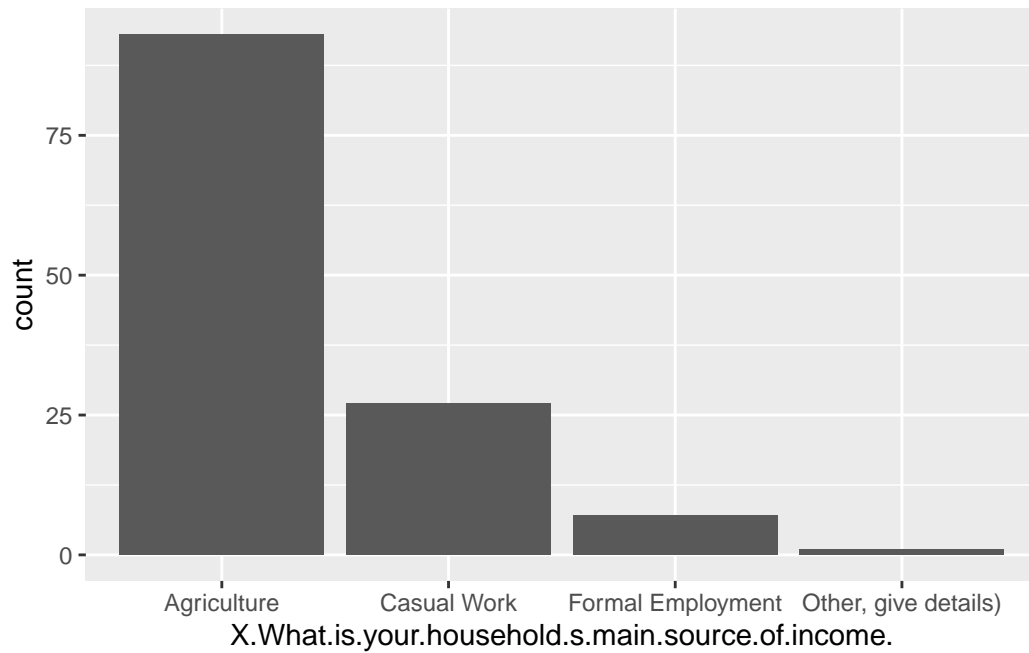
```
# A tibble: 1 x 8
  n_missing numeric.mean numeric.sd numeric.p0 numeric.p25 numeric.p50
  <int>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1      11      88.4      88.0         2        30        60
  numeric.p75 numeric.p100
  <dbl>      <dbl>
1     120      480
```

```
# A tibble: 1 x 8
  n_missing numeric.mean numeric.sd numeric.p0 numeric.p25 numeric.p50
  <int>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1         1      34.7      43.1         3        10        15
  numeric.p75 numeric.p100
  <dbl>      <dbl>
1        35       300
```

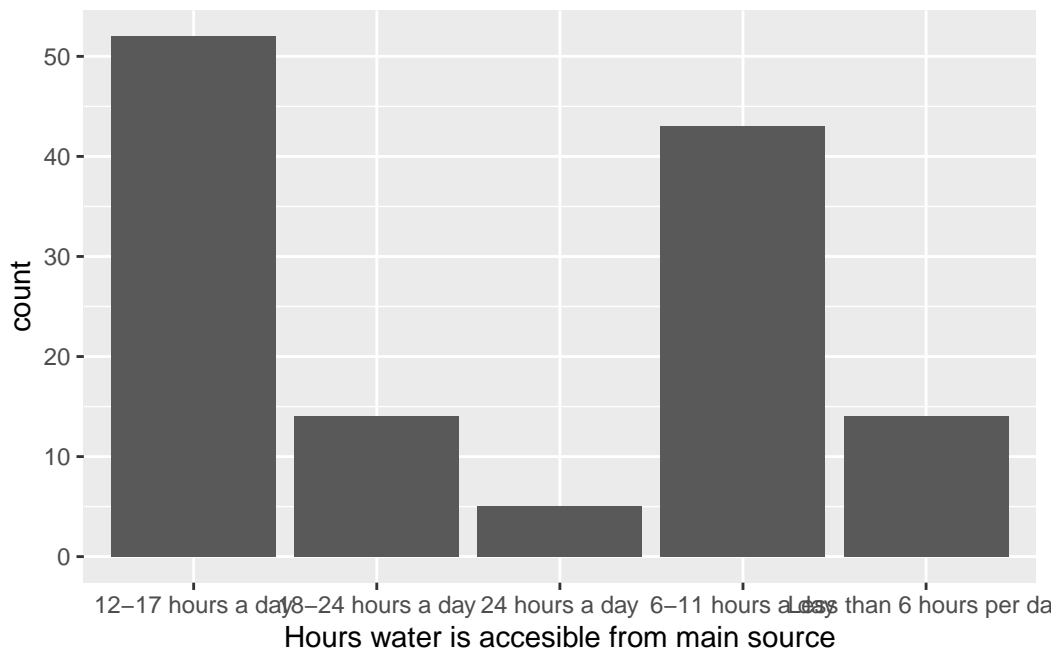
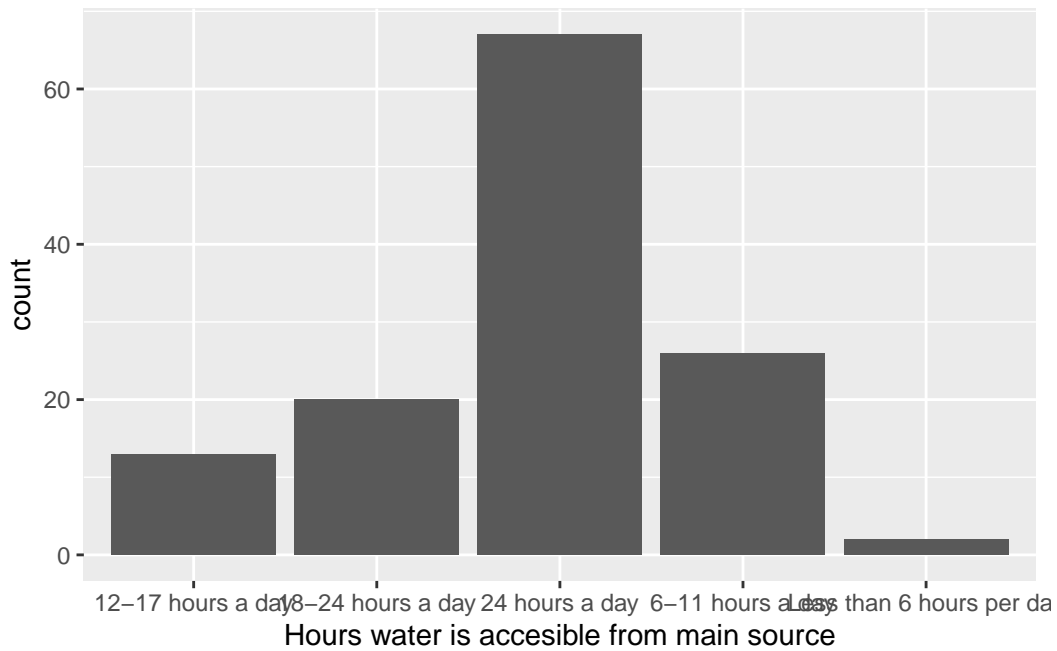
Both datasets are clearly right skewed therefore using the median amount of time for water collection drastically decreased from the baseline to the recent project survey. The median amount of time for water collection decreased from about 60 to 15 minutes. With the ranges also being drastically smaller for the minutes for the post survey.



Based on both the graphs, there is no overlap in the waterpoint site names. Having a project and baseline, wondering if there is any other data that would help keep one dataset as a base for comparison.

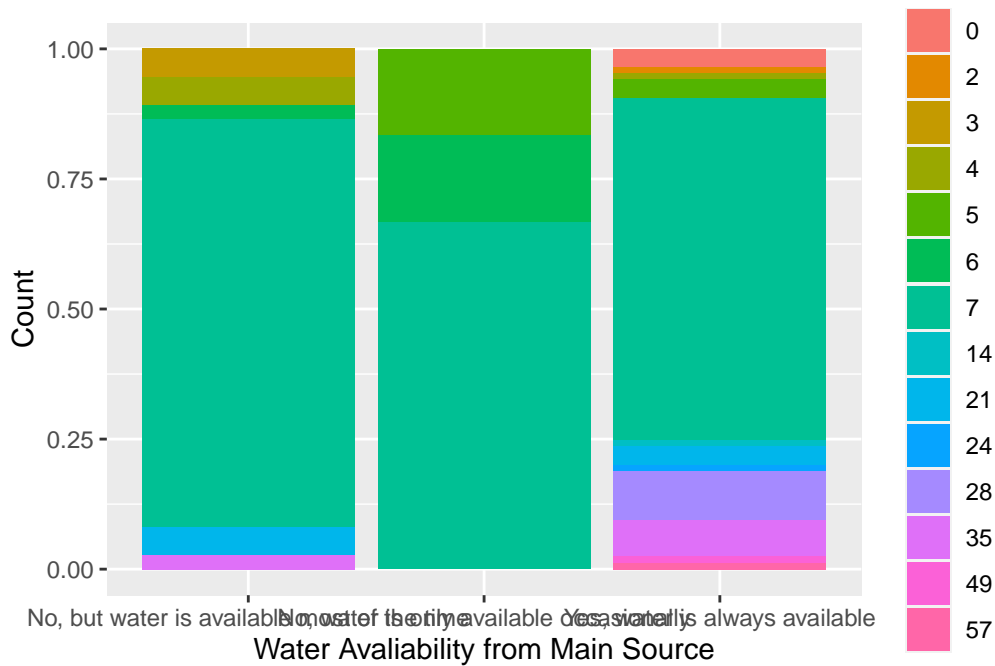


Both have similar household income levels, this could be used a baseline of comparison rather than the waterpoint site name.

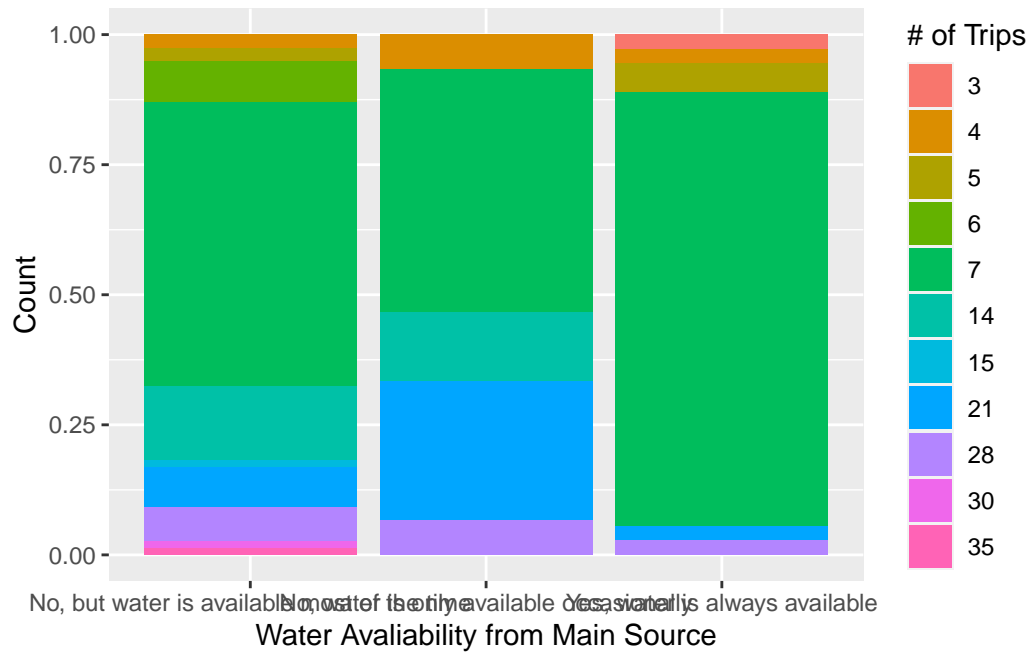


Based on both the graphs, there is a increase in the “12-17 hours a day” water availability post the baseline survey. However, an interesting thing that the 24 hours a day drastically decreased from around 65 to 5 wells available.

## Bi-variate EDA



For the instances that water is always available, there seems to be a higher percentage of more trips taken to the main water source. One thing that did stand out to me was the response of “No, but water is available most of the time” had the most amount of “medium” trips taken (around 5-6 trips). Assuming that they had bigger containers to store and carry water, therefore needing less trips. The people who have constant water access, have smaller containers and therefore make more trips in order to obtain water from their main water source.



We see a drastic difference in the “Yes, water is always available” column as the number of trips frequented dropped from a high of 24-57 trips to a majority of only 6-7 trips. Assuming the project survey was done months after the wells were drilled, we could see that it helped in reducing the trips made by local residents.